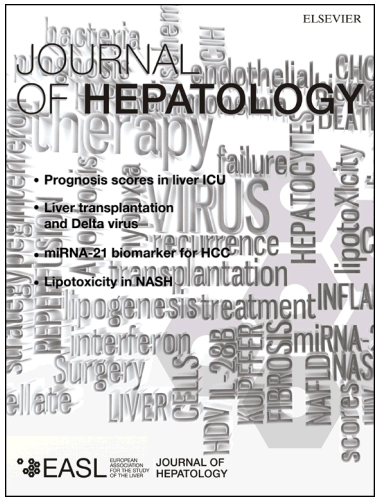
Applicability and prognostic value of histologic scoring systems in primary sclerosing cholangitis

De Vries, Elisabeth M.g.; Verheij, Joanne; Hubscher, Stefan G.; Leeflang, Mariska M.g.; Boonstra, Kirsten; Beuers, Ulrich; Ponsioen, Cyriel Y.

DOI:
10.1016/j.jhep.2015.06.008

Accepted Manuscript



Applicability and Prognostic Value of Histologic Scoring Systems in Primary Sclerosing Cholangitis

Elisabeth M.G. de Vries, Joanne Verheij, Stefan G. Hubscher, Mariska M.R. Leeflang, Kirsten Boonstra, Ulrich Beuers, Cyriel Y. Ponsioen

PII: S0168-8278(15)00399-2
DOI: http://dx.doi.org/10.1016/j.jhep.2015.06.008
Reference: JHEPAT 5712

To appear in: *Journal of Hepatology*

Received Date: 13 February 2015
Revised Date: 13 May 2015
Accepted Date: 8 June 2015

Please cite this article as: de Vries, E.M.G., Verheij, J., Hubscher, S.G., Leeflang, M.M.R., Boonstra, K., Beuers, U., Ponsioen, C.Y., Applicability and Prognostic Value of Histologic Scoring Systems in Primary Sclerosing Cholangitis, *Journal of Hepatology* (2015), doi: http://dx.doi.org/10.1016/j.jhep.2015.06.008

Applicability and Prognostic Value of Histologic Scoring Systems in Primary Sclerosing Cholangitis

Elisabeth MG de Vries[1], Joanne Verheij[2], Stefan G Hubscher[3], Mariska MR Leeflang[4], Kirsten Boonstra[1], Ulrich Beuers[1], Cyriel Y Ponsioen[1]

[1] Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, the Netherlands

[2] Department of Pathology, Academic Medical Center, Amsterdam, the Netherlands

[3] School of Cancer Sciences, University of Birmingham and Department of Cellular Pathology, Queen Elizabeth Hospital, Birmingham, United Kingdom

[4] Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam, the Netherlands

Correspondence

C.Y. Ponsioen, MD PhD

Dept. of Gastroenterology and Hepatology

Academic Medical Center

Meibergdreef 9,1105 AZ,

Amsterdam, The Netherlands

Fax nr: 0031206917033, phone nr: 0031205666012

Email: c.y.ponsioen@amc.uva.nl

Electric word count manuscript:

5595

Number of tables:

4

Number of figures:

2

List of abbreviations

PSC = primary sclerosing cholangitis

LTx = liver transplantation

CRS = cirrhosis related symptoms

CI = confidence interval

MRCP = magnetic resonance cholangiopancreatography

ERCP = endoscopic retrograde cholangiopancreatography

AIH = auto-immune hepatitis

CBP = copper binding protein

PBC = primary biliary cirrhosis

IQR = inter quartile range

AST = aspartate aminotransferase

ALT = alanine aminotransferase

ALP = alkaline phosphatase

γGT = gamma-glutamyl transferase

HE = haematoxylin and eosin

CA = cholangitis activity

HA = hepatitis activity

HR = hazard ratio

xULN = times upper limit of normal

IQR = Inter quartile range

r = correlation coefficient

UDCA = ursodeoxycholic acid

Key words: Primary sclerosing cholangitis; histology; histologic scoring system; prognosis; surrogate endpoint.

Conflicts of Interest and Source of Funding:

All authors disclosed no financial relationship relevant to this publication.

Authors contribution

CP designed the study and supervised the project. EdV collected patient data and histologic material, performed the statistical analyses, interpretation of the data and prepared the first draft of the manuscript. ML supervised the statistical analyses. JV and SH scored the biopsies. KB identified PSC patients and collected patient data. All authors reviewed the manuscript for critical content, and approved the final version.

Abstract

Background Primary sclerosing cholangitis (PSC) is a chronic cholestatic liver disease. At present there is no appropriate histologic scoring system available for PSC, evaluating both degree of necroinflammatory activity (grade) and fibrosis (stage). Aim of this study was to assess if three scoring systems, commonly used in different liver diseases could be applied for grading and/or staging of PSC.

Methods Sixty-four PSC patients from a Dutch cohort, who underwent diagnostic liver biopsy, were included. Staging was scored using Ishak, Nakanuma, and Ludwig systems. Grading was scored using Ishak and Nakanuma systems. Three measures of outcome were defined; transplant-free survival, time to liver transplantation (LTx) and occurrence of cirrhosis related symptoms (CRS). Association of grade and stage with outcome was estimated using Kaplan Meier-log-rank test, and Cox regression analysis. Correlation with biochemistry was assessed by Spearman's rank test.

Results There were strong associations between disease stage measured by Ishak, Nakanuma and Ludwig staging systems with both outcome measure transplant-free survival (Hazard ratio (HR)2.56; 95%CI 1.11-5.89, HR6.53; 95%CI 2.01-21.22, HR1.94; 95%CI 1.00-3.79, respectively), and time to LTx (HR4.18; 95%CI 1.51-11.56, HR7.05; 95%CI 1.77-28.11, HR3.13; 95%CI 1.42-6.87, respectively). Ishak and Nakanuma grading systems were not associated with CRS. Weak correlations between histopathology and liver biochemistry were shown.

Conclusion Applying the Nakanuma, Ishak, and Ludwig histopathological staging systems is feasible and clinically relevant given their association with transplant-free survival and time to LTx. This suggests these staging systems may be likely candidates for surrogate endpoints and stratification purposes in clinical trials in PSC.

Word count: 249

INTRODUCTION

Primary sclerosing cholangitis (PSC) is a chronic cholestatic liver disease, characterized by progressive bile duct scarring and destruction, leading to biliary fibrosis and eventually progression to end-stage liver cirrhosis.[1] PSC diagnosis is established by means of cholangiography, performed by magnetic resonance cholangiopancreatography (MRCP) which is implemented as “golden standard”, and has replaced the more invasive endoscopic retrograde cholangiopancreatography (ERCP).[2,3] In case of large duct PSC, cholangiogram typically shows biliary strictures interchanged with dilatations creating the “beaded” appearance, and routine liver biopsy is not necessary to confirm diagnosis.[4] However, more subtle caliber changes can easily be missed on MRCP, and conversely, due to the moderate resolution of MRCP, false-positive findings may occur. In case of doubt, suspicion of small duct PSC, or auto-immune hepatitis (AIH) overlap syndrome, liver biopsy is indicated and essential to confirm diagnosis.[5,6] Furthermore liver biopsies have been used for the evaluation of treatment efficacy in therapeutic trials.[7–13]

Histologic changes seen in PSC include the characteristic periductal concentric fibrosis leading to bile duct obliteration, infiltration of inflammatory cells in portal tracts, loss of bile ducts, bile ductular reaction and focal accumulation of copper binding protein (CBP).[1,14,15] Using liver histology, disease severity and progression can be assessed in terms of grade and stage.[16] Grade is usually used to describe the degree of necroinflammatory activity as measurement of the severity of the underlying disease process, while stage generally reflects the degree of fibrosis and cirrhosis as measurement of disease progression.[16]

If a PSC appropriate histologic scoring system would have prognostic significance in terms of predicting the occurrence of solid clinical endpoints, liver

histology may be an important candidate for the evaluation of treatment efficacy in therapeutic trials. Evaluation of treatment efficacy by solid clinical endpoints such as death or liver transplantation (LTx), is hindered by the chronic disease course and the low prevalence of PSC.[18] Therefore, surrogate endpoints, including clinical biomarkers, biochemical biomarkers, Mayo risk score, and liver histology have often been used, but never been validated.[7–13] Currently, the lack of properly validated surrogate endpoints for clinical trials is one of the major challenges in PSC research. An important asset of liver histology is its face-validity, meaning that liver biopsy directly measures the degree of disease severity in the affected organ. However, it is currently unknown if face-validity is maintained in PSC livers where the patchy distribution may give rise to confounding sampling variability.[19]

At present there is no specific PSC histologic scoring system with clinical significance, to evaluate both disease grade and stage. Commonly, the Ludwig and Ishak systems have been used to grade and stage histologic disease severity in PSC.[17,20] A drawback of the Ludwig staging system is that it was designed primarily to assess disease progression of primary biliary cirrhosis (PBC).[17] Furthermore, the Ludwig system does not separately score disease grade, and instead incorporates features such as portal and periportal inflammation, which are probably better regarded as manifestation of disease grade rather than stage.

Recently, Nakanuma et al, have proposed a new grading and staging system for primary biliary cirrhosis (PBC), which takes into account particular features that are shared with PSC, such as the presence of copper binding protein and loss of bile ducts.[21]

The aim of this study was to determine the prognostic value of three different scoring systems, designed primarily to assess disease grade and/or stage in chronic

hepatitis (Ishak *et al.*) or PBC (Ludwig *et al.*, Nakanuma *et al.*), for grading and/or staging of PSC.[17,20–22]

PATIENTS AND METHODS

Study design

This cohort study is part of the 'Epi PSC PBC project', a large population-based cohort study of PSC and PBC in the Netherlands. All PSC patients alive on January 1st 2000 and living in a geographically defined area of 6 adjacent provinces comprising 50% of the Dutch population were included in this study between January 1st 2008 and December 31st 2011. The case-finding and case-ascertainment methods have been described previously.[18] The protocol was approved by the central Committee for Research Ethics in Utrecht and all 44 local ethics committees of the participating hospitals in the Netherlands (trialregister.nl number, NTR2813).

PSC diagnosis was based on: 1) elevated alkaline phosphatase and gamma-glutamyltransferase, not explained otherwise, 2) presence of characteristic bile duct changes with multifocal strictures and segmental dilatations on ERC or magnetic resonance cholangiography (MRC) and/or 3) liver histology and 4) no evidence for secondary sclerosing cholangitis. When criteria 1, 3 and 4 were fulfilled in the absence of cholangiographic abnormalities on MRC or ERC, cases were diagnosed as small duct PSC.[6] Autoimmune hepatitis (AIH) overlap syndrome (PSC- AIH) is ill defined. A diagnosis of PSC-AIH was made in patients with a characteristic cholangiogram who, in addition, met the simplified AIH criteria.[23]

PSC patients from the Epi PSC PBC cohort, who underwent diagnostic liver biopsy or liver biopsy to assess disease severity at time of diagnosis between 1978 and 2011, were included. Patients with PSC-AIH overlap syndrome were excluded. Biochemical values of aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), gamma-glutamyl transferase (γGT) and total

bilirubin at time of liver biopsy (range two months before until one month after liver biopsy) were retrieved from hospital databases and Mayo risk score was calculated. Clinical data reflecting liver cirrhosis related symptoms (CRS) at follow up were collected from patient files. CRS included gastro-esophageal varices and variceal bleeding, both assessed by gastrointestinal endoscopy, ascites, and splenomegaly assessed clinically and/or by imaging, and hepatic encephalopathy.

Tissue preparation and histologic evaluation

Original liver specimens, fixed in formalin and embedded in paraffin, as well as original liver stained sections were collected from the pathology department diagnostic archives. From each paraffin block thin sections were cut for haematoxylin and eosin (HE), connective tissue (Sirius red) and orcein stainings. Orcein staining was used to assess degree of copper binding protein (CBP) deposition in hepatocytes. If well preserved and available, original stained sections were used for histologic evaluation of biopsies. Otherwise, new stains were carried out. Grade and stage of biopsy specimens were evaluated using the three systems referred to above by two expert liver pathologists (JV & SH) in tandem using a multihead microscope, with the intention to reach consensus.

Grading

Grading was scored according to the Ishak system, evaluating degree of interface hepatitis (score 0-4), confluent necrosis (score 0-6), lobular inflammation (score 0-4) and portal inflammation (score 0-4).[20] Furthermore slides were scored according to the Nakanuma system, encompassing degree of cholangitis activity (CA) (score 0-3) and hepatitis activity (HA)(score 0-3).[21] (Supplementary appendix, table 1.)

Staging

Staging was performed according the Ishak, as well as the Nakanuma and Ludwig systems.[17,20,21] With Ishak staging system degree of fibrosis is evaluated (0-6). The Nakanuma staging system is based on semi-quantitative scoring of three histological features - fibrosis (score 0-3), bile duct loss (score 0-3) and deposition of orcein positive granules (score 0-3). The final Nakanuma stage is obtained from the total score of these three features. Stage I (no or minimal progression) is a score of 0, stage II (mild progression) a score of 1-3, stage III (moderate progression) a score of 4-6 and stage IV (advanced progression) a score of 7-9.[24] Ludwig staging system consists of 4 stages; stage I, cholangitis or portal hepatitis; stage II, periportal fibrosis or hepatitis; stage III, septal fibrosis, bridging necrosis or both; and stage IV, biliary cirrhosis. (Supplementary appendix, table 2.)

Endpoints

For analyses of association with endpoints, three different endpoints were chosen. The first endpoint was transplant-free survival, defined as time to PSC-related death (death from end-stage liver disease, liver surgery, cholangiocarcinoma and colorectal carcinoma), LTx and presentation with cholangiocarcinoma. Since the occurrence of cholangiocarcinoma may not be predictable by liver histology at time of diagnosis, second endpoint was time to LTx alone. The third endpoint was the occurrence of liver cirrhosis related symptoms (CRS) at follow-up.

Statistical Analysis

Patient characteristics were expressed as mean ± standard deviation and median and interquartile range where appropriate.

Association of histologic grade and stage with transplant-free survival, time to LTx and development of CRS was estimated using Kaplan Meier survival curve and Wilcoxon log-rank test. Due to relatively small sample size, and the resulting small amount of patients per grade and stage, survival analyses were performed in grouped subcategories. In this reclassification the order of severity of grade/stage was maintained; those subgroups with very few or no patients were grouped together with the grade/stage of most similar severity. Ishak grading components interface hepatitis, lobular inflammation and portal inflammation were reclassified in score 0, 1, ≥2 (original score: 0-4), component confluent necrosis was reclassified in 0, ≥1 (original score: 0-6), and the total Ishak grade was reclassified in score 1-4 (original score: 0-1 = 1; 2,3 = 2; 4 = 3; ≥5 = 4). Nakanuma grading component cholangitis activity was reclassified in 0, 1, ≥2 (original score: 0-3), component hepatitis activity in 0, ≥1 (original score: 0-3), and the total Nakanuma grade score in 0, 1, ≥2 (original score: 0-6). Ishak stage was reclassified in score 1-3 (original score : 0,1 = 1; 2,3 = 2; ≥4 = 3). For the Nakanuma staging system the original scoring system was maintained. Ludwig staging system was scored 1-4 in which a Ludwig score of 0 was classified as score 1.

Cox proportional hazard analysis was performed and associated hazard ratio (HR) was calculated for clinical, biochemical and histopathological variables. Exploratory analyses for correlations between liver biochemistry and histologic grade and stage to assess if liver biochemistry could reflect the degree of liver injury measured by liver histology, were performed by Spearman's ranking test.

Statistical analyses were performed using SPSS version 20.0 software (SPSS, Inc., Chicago, IL). $p<0.05$ was considered statistically significant.

RESULTS

Patient and biopsy characteristics

Sixty-four patients were included, 40 male and 24 female, with a median follow-up of 112 months (IQR 70-172). Patient characteristics, serum liver tests, and characteristic histologic features are summarized in table 1. Laboratory results are expressed in times upper limit of normal (xULN), as assays varied between hospitals and over time. Biopsies had a median length of 14 mm (IQR 11-19). The median number of portal tracts was 13 (IQR 9-19). Median disease duration at time of biopsy was 0 months (range 0-20).

Distribution grade and stage

Consensus on the evaluation of grade and stage was reached in 100% of cases.

Grading:

Ishak grading component interface hepatitis showed a median grade of 1 (range 1-2). Confluent necrosis was a relative infrequent finding in PSC liver biopsies with a median grade of 0 (range 0-1). The median grade of lobular inflammation and portal inflammation was 1 (range 0-2) and 1 (range 0-3), respectively. The majority of liver biopsies did not show Nakanuma grading components CA with a median of 0 (range 0-3) nor HA, median 0 (range 0-2). Distribution of total Ishak and Nakanuma grade is shown in figure 1A,B.

Staging:

Figure 1C represents the Ishak stage, with a median of 2 (range 0-6). In figure 1D the Nakanuma stage (median 2 (range1-4)) is shown, derived from the components

fibrosis, bile duct loss and deposition of orcein positive granules score. Ludwig stage is displayed in figure 1E, showing a normal distribution (mean 2, SD 1).

The distribution of Ishak and Nakanuma grading components, and the Nakanuma staging components is illustrated in figure 1 of the supplementary appendix.

Patient outcome

PSC related death, liver transplantation and cholangiocarcinoma

Long-term follow up data were present for all patients. A total of 11 (17%) patients developed an endpoint; 3 patients were diagnosed with cholangiocarcinoma of whom 2 patients died and eight patients underwent LTx. The median time to endpoint was 86 months (IQR 43-142).

Liver cirrhosis related symptoms:

Data on the occurrence of CRS were available for 63 patients. A total of 16 (25%) patients developed CRS, of which nine developed more than one CRS. Nine (14%) patients developed varices, of whom one had a variceal bleeding. Ascites was found in six (10%) patients, splenomegaly in twelve (19%) and three (5%) patients presented with hepatic encephalopathy. The median time until presentation of the first CRS was 53 months (range 0-280), one patient presented with CRS at time of liver biopsy.

Association of histologic grade and stage with patient outcome

The prognostic significances of clinical, biochemical and histopathological parameters in predicting transplant-free survival, time to LTx and CRS, calculated by univariable Cox proportional hazard analyses are summarized in table 2.

Grading

No significant association of the Ishak and Nakanuma grading systems and outcome measures transplant-free survival, time to LTx and CRS were shown.

Staging

A significant association was shown for Ishak staging system and transplant-free survival (p=0.04) as well as Ishak stage and time to LTx (p=0.005) (figure 2A,B). Univariable Cox proportional hazard analysis revealed a hazard ratio of 2.56; 95% CI 1.11-5.89 (p=0.03) and HR 4.18; 95% CI 1.51-11.56 (p=0.006) for transplant-free survival and time to LTx, respectively (table 2). Nakanuma staging components fibrosis and CBP deposition were significantly associated with transplant-free survival and time to LTx, in which degree of CBP deposition was most pronounced (p <0.001 vs p=0.005 respectively (supplementary appendix, table 3). This resulted in a strong significant association of Nakanuma staging system as a whole with both transplant-free survival (p=<0.001) and time to LTx (p<0.001) (figure 2 C,D). The accompanying hazard ratio was 6.53; 95%CI 2.01-21.22 (p=0.002) and HR 7.05; 95% CI 1.77-28.11 (p=0.006) for transplant-free survival and time to LTx, respectively. (table 2). The Ludwig staging system also showed a significant association with outcome measurements transplant-free survival and time to LTx (log-rank p=0.027 and 0.002, respectively), HR 1.94; 95% CI 1.00-3.79 (p=0.05) and HR 3.13; 95% CI 1.42-6.87 (p= 0.005), respectively. (Figure 2 E,F; table 2)

The endpoint cirrhosis related symptoms showed a significant association with the Ludwig staging system (log rank p=0.044). There was no association between the Nakanuma staging system as a whole and the development of CRS, though the Nakanuma staging component fibrosis did show a significant association (log rank

p=0.04). The Ishak staging system was not associated with the development of cirrhosis related symptoms (supplementary appendix, table 3).

Univariable analyses of the other parameters showed that bilirubin and ALP were predictive factors for time to LTx. Though a trend for similar associations was seen for the endpoint transplant-free survival, this was not significant. However, MRS was a predictive factor for transplant-free survival. (table 2).

An overview of median survival time per grade and stage, as well as the associations of histologic stage and grade with the three outcome measurements, calculated by log-rank test is given in the supplementary appendix, table 3.

Correlation between histologic grade and stage and liver biochemistry values

In table 3a and 3b correlation between liver biochemistry and histologic grade and stage are shown, calculated using Spearman's rank correlation test. For some patients serum liver tests at time of liver biopsy were not available. Liver biochemistry could be retrieved for AST n=56, ALT n= 51, ALP n=52, γGT n=46, total bilirubin n= 49.

For grading, significant positive correlations between serum ALP level and the Ishak grading component portal inflammation (r=0.31, *p*=0.027), Ishak grading system as a whole (r=0.31, p=0.023) and the Nakanuma grading component HA (r=0.28, *p*=0.047) were shown.

For staging, a positive correlation between total bilirubin level and Ishak staging system (r=0.32, *p*=0.027), as well as Nakanuma staging component fibrosis were demonstrated (r=0.29, *p*=0.042).

Correlation coefficients did not exceed 0.33.

DISCUSSION

With this study we have demonstrated that the Ishak, Nakanuma and Ludwig histological *staging* systems are predictive indicators of transplant-free survival and time to LTx in PSC.

The Nakanuma staging system appeared to have the strongest predictive power, given its highest incremental hazard of long-term outcome corresponding to the stage progression. This may be explained by the fact that the Nakanuma staging system includes features that could be considered PSC appropriate. When focusing on the prognostic value of these individual components, degree of fibrosis and CBP deposition both predicted transplant-free survival and time to LTx, in which degree of CBP deposition was most discriminative ($p < 0.001$, $p=0.005$ respectively). The Nakanuma staging component bile duct loss was not associated with outcome. This is in contrast with findings in PBC, where bile duct loss has been reported to be a predictor of disease progression.[25] Moreover, in PSC a peripheral liver biopsy might not be the best way to evaluate the extent of bile duct loss, since the disease affects the entire biliary tree and obliteration or obstruction of the larger central bile ducts may greatly affect progression to liver failure.[15]

The prognostic value of the Ishak and Ludwig staging system, as well as Nakanuma component fibrosis implicate that degree of fibrosis is of important prognostic value in PSC. Similar findings have recently been demonstrated by Ruiz e*t al*., who analyzed radiologic disease course in PSC by 3-dimensional MRC, and found that risk of progression was mainly dependent on the effects of biliary disease on liver parenchyma, rather than the effects of severe stricturing. [26]

Necroinflammatory activity may predispose to development of CRS, however, no associations of histologic *grading* systems with CRS were shown. In PBC,

associations between histology and CRS seem to be more pronounced. Kakuda *et al.* showed a significant prognostic value of Nakanuma components fibrosis and CBP deposition, as well as Nakanuma and Ludwig staging systems as a whole for the development of cirrhosis related conditions before ursodeoxycholic acid (UDCA) treatment.[24] In a retrospective cohort of 58 PBC patients Chan *et al*. confirmed the Nakanuma system as a prognostic factor for liver related events, while the Ludwig system was not.[27] It must be stated that, despite that PSC and PBC are both cholangiopathies, sharing biochemical, clinical and a some morphological features, considerable differences exist in both disease entities[15]. The patchy distribution of affected bile ducts throughout the liver, with resulting sampling variability may be greater in PSC than in PBC.[19] Furthermore, the differences in statistical analyses as well as in definition of CRS between studies may contribute to these conflicting results.

The extent to which histologic features, scored by the different scoring systems were present in PSC biopsies may indicate their applicability in PSC. The Nakanuma grading component cholangitis activity – though well recognized in PBC – is not typically seen in PSC. This was reflected in the presence of cholangitis activity in only 38% of cases in our cohort and the associated lack of predictive value. Hepatitis activity was detected in a minority of 27% of cases and had no predictive value. This may partly be explained by the exclusion of patients with auto-immune hepatitis overlap syndrome, in whom more severe hepatitis activity would be expected. PSC appropriate Nakanuma staging components bile duct loss and orcein positive granules were present in only 50% of cases. This could be due to the inclusion of biopsies taken at time of diagnosis, creating a cohort of patients with relatively early-stage disease.

Exploratory analyses for correlations between liver biochemistry and histologic grade and stage were performed to assess if liver biochemistry could reflect the degree of liver injury measured by liver histology. No strong correlations could be found, since correlation coefficients did not exceed 0.33. This suggests that liver biochemistry does not adequately reflect the degree of histologic injury. This finding is supported by a recent study of Queen *et al.* who evaluated the clinical course, endoscopic and pathologic findings of patients with normal liver biochemistry and showed that PSC patients can have cirrhosis and significant ductal disease, in the setting of normal liver biochemistry.[28]

In 1995, Olsson *et al.* demonstrated a sampling variability of at least 20% for blind needle biopsy in PSC, which is often thought to refute its use as a prognostic measure.[29] The results of our study show that sampling variability is apparently not a major confounder. This is in line with results from a study by Angulo *et al.,* who assessed the time course of histologic progression in PSC.[30] They observed progression in Ludwig stage in liver biopsies over two years' time in 53% of PSC patients with an initial Ludwig stage of I-III.[30] In a Markov model they estimated that the risk of progression from Ludwig stage II to stage III of IV was 66% after two years. These results indicate that despite the issue of sampling variability, progression of liver disease can be evaluated using liver biopsy.[30] In addition, several clinical trials have included histologic change in grade-/stage as clinical endpoint.[7–13] Angulo et al. noted a significant improvement in the degree of portal inflammation – grade – after one year of budesonide treatment.[13] Degree of fibrosis and disease stage progressed despite treatment.[13] In a 2-year double-blind placebo-controlled study of high-dose UDCA, Mitchell et al. showed a significant reduction in progression of disease stage in the treatment group.[9] Fifty percent of

the placebo group showed no change in disease stage, while the other fifty percent showed progression by a single stage.[9] These studies indicate that both progression and improvement of liver disease grade-/stage can adequately be assessed by liver biopsy. [9,13,30] In addition, with the present study the prognostic value of liver biopsy on outcome was demonstrated. Therefore, liver histology appears to be a useful biomarker for progression, and its use as outcome measure in clinical trials deserves consideration. Moreover, for regulatory bodies, histology is still the gold standard as outcome parameter in many liver diseases.

A limitation of this study is that multivariable Cox proportional hazard regression analysis, to analyze if the prognostic value of the histology is independent of clinical and biochemical variables, could not be applied for methodological reasons. When performing a multivariable Cox proportional hazard regression analysis, the total numbers of events – not the total number of included cases – is important, to retain the validity of the outcome. [31] The most appropriate number of events per variable (EPV) to use in a multivariable analysis is 10. [32,33] For the present study a total of 140 events – an increase of the sample size up to 10 times – would be necessary to generate stable and interpretable results by multivariable analysis, which was unfortunately not feasible.

The retrospective collection of serum liver tests and data on occurrence of liver CRS may be incomplete and hence may have led to a relative underestimation of the occurrence of liver CRS. However, follow-up data on solid clinical endpoints were complete in all 64 patients. Its retrospective character is also an asset in the sense that it yielded a median follow-up time of 112 months. Guido *et al.* described liver biopsy samples of 22 mm or more in length, with at least 11 portal tracks to be most reliable for grading and staging at least in the setting of viral hepatitis.[34] Although

the quality of liver biopsies of the present study was high with a median amount of portal tracks of 13 (IQR 9-19) and a median biopsy length of 14 mm (IQR 11-19), these optimal requirements were not met in all cases. The biopsies included in this study were all taken at time of diagnosis, which may result in spectrum bias. However, in figure 1. a normal distribution of grades and stages for most of the classification systems is shown, demonstrating that PSC patients diagnosed in both early and late disease stage were included in this study. Lastly the percentage of small duct PSC in this study (16%) is relatively high in comparison with the 9% small duct PSC patients reported in a large population based epidemiology study.[18] Small duct PSC patients are known to have a better prognosis, which could influence results of outcome. For this reason analyses for associations between histological grading and staging and outcome were repeated in only the large duct patients of this cohort. Similar results were demonstrated (data not shown).

In conclusion, applying the Nakanuma, Ishak, and Ludwig histopathological staging systems in PSC is feasible and clinically relevant. Our results support the notion that histopathological scoring systems are likely candidates for surrogate endpoints and stratification in clinical trials in PSC, and may be used to assess noninvasive biomarkers for future trials. Validation of the applicability of these staging systems as well as determination of inter-observer variability to test the robustness of these scoring systems in PSC in a large multi-center cohort is warranted.

ACKNOWLEDGEMENTS

The Epi PSC PBC study was supported by unrestricted grants from Dr. Falk Benelux, AbbVie and Tramedico.

We thank all participating pathology laboratories for providing liver histology material.

REFERENCES

[1] Hirschfield GM, Karlsen TH, Lindor KD, Adams DH. Primary sclerosing cholangitis. Lancet 2013;382:1587–99.

[2] Maneesh D, Elmunzer BJ, Dwamena B, Higgins P. Primary Sclerosing Cholangitis: Meta-Analysis of Diagnostic Performance of MR Cholangiopancreatography. Radiology 2010;256:387–96.

[3] Weber C, Kuhlencordt R, Grotelueschen R, Wedegaertner U, Ang TL, Adam G, et al. Magnetic resonance cholangiopancreatography in the diagnosis of primary sclerosing cholangitis. Endoscopy 2008;40:739–45.

[4] Maccarty R, Larusso N, Wiesner R, Ludwig J. Primary Sclerosing Cholangitis: Findings on Cholangiography and Pancreatography. Radiology 1983;149:39–44.

[5] Chapman R, Fevery J, Kalloo A, Nagorney DM, Boberg KM, Shneider B, et al. Diagnosis and management of primary sclerosing cholangitis. AASLD Practice Guidelines. Hepatology 2010;51:660–78.

[6] EASL. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. J Hepatol 2009;51:237–67.

[7] Lindor KD. Ursodiol for primary sclerosing cholangitis. NEJM 1997;336:691–5.

[8] Lindor KD, Kowdley K V, Luketic V a C, Harrison ME, McCashland T, Befeler AS, et al. High-dose ursodeoxycholic acid for the treatment of primary sclerosing cholangitis. Hepatology 2009;50:808–14.

[9] Mitchell S a., Bansi DS, Hunt N, Bergmann K Von, Fleming K a., Chapman RW. A preliminary trial of high-dose ursodeoxycholic acid in primary sclerosing cholangitis. Gastroenterology 2001;121:900–7.

[10] Färkkilä M, Karvonen A-L, Nurmi H, Nuutinen H, Taavitsainen M, Pikkarainen P, et al. Metronidazole and ursodeoxycholic acid for primary sclerosing cholangitis: a randomized placebo-controlled trial. Hepatology 2004;40:1379–86.

[11] Hommes DW, Erkelens W, Ponsioen C, Stokkers P, Rauws E, van der Spek M, et al. A double-blind, placebo-controlled, randomized study of infliximab in primary sclerosing cholangitis. J Clin Gastroenterol 2008;42:522–6.

[12] Thiel DH Van, Carroll P, Abu-elmagd K, Irish W, Sc M, Mcmichael J, et al. Tacrolimus (FK), a Treatment for Primary Sclerosing Cholangitis; Results of an Open-Label Preliminary Trial. Am J Gastroenterol 2010;90:455–9.

[13] Angulo P, Batts KP, Jorgensen R a, LaRusso N a, Lindor KD. Oral budesonide in the treatment of primary sclerosing cholangitis. Am J Gastroenterol 2000;95:2333–7.

[14] Barbatis C, Grases P, Sheperd H, Chapman R, Trowlell J, Jewell DPJ, et al. Histological features of sclerosing cholangitis in patients with chronic ulcerative colitis. J Clin Pathol 1985;38:778–83.

[15] Scheuer P, Path F. Pathologic Features and Evolution of Primary Biliary Cirrohosis and Primary Sclerosing Cholangitis. Ludwig Symposium on Biliary Disorders-PartII. Mayo Clin Proc 1998;73:179–83.

[16] Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. J Hepatol 2007;47:598–607.

[17] Ludwig J, Dickson E, McDonald G. Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). Virchows Arch A Pathol Anat Histol 1978;22:103–12.

[18] Boonstra K, Weersma RK, van Erpecum KJ, Rauws E a, Spanier BWM, Poen AC, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. Hepatology 2013;58:2045–55.

[19] Portmann B, Zen Y. Inflammatory disease of the bile ducts-cholangiopathies: liver biopsy challenge and clinicopathological correlation. Histopathology 2012;60:236–48.

[20] Ishak K, Baptista a, Bianchi L, Callea F, De Groote J, Gudat F, et al. Histological grading and staging of chronic hepatitis. J Hepatol 1995;22:696–9.

[21] Nakanuma Y, Zen Y, Hararda K, Sasaki M. Application of a new histological staging and grading system for primary biliary cirrhosis to liver biopsy specimens: Interobserver agreement. Pathol Int 2010;60:167–74.

[22] Hiramatsu K, Aoyama H, Zen Y, Aishima S. Proposal of a new staging and grading system of the liver for primary biliary cirrhosis. Histopathology 2006;49:466–78.

[23] Hennes EM, Zeniya M, Czaja AJ, Parés A, Dalekos GN, Krawitt EL, et al. Simplified criteria for the diagnosis of autoimmune hepatitis. Hepatology 2008;48:169–76.

[24] Kakuda Y, Hararda K, Sawada-Kitamura S, Ikeda H, Nakanuma Y. Evaluation of a new histologic staging and grading sustem for primary biliary cirrhisis in comparison with classical systems. Hum Pathol 2013;44:1109–17.

[25] Kumagi T, Guindi M, Fischer SE, Arenovich T, Abdalian R, Coltescu C, et al. Baseline Ductopenia and Treatment Response Predict Long-Term Histological Progression in Primary Biliary Cirrhosis. Am J Gastroenterol 2010;105:2186–94.

[26] Ruiz A, Lemoinne S, Carrat F, Corpechot C, Chazouillères O, Arrivé L. Radiologic course of primary sclerosing cholangitis: assessment by three-dimensional magnetic resonance cholangiography and predictive features of progression. Hepatology 2014;59:242–50.

[27] Chan AWH, Chan RCK, Wong GLH, Wong VWS, Choi PCL, Chan HLY, et al. Evaluation of histological staging systems for primary biliary cirrhosis: correlation with clinical and biochemical factors and significance of pathological parameters in prognostication. Histopathology 2014;65:174–86.

[28] Queen T, Cox K, Adler DG. Primary Sclerosing Cholangitis in the Setting of Normal Liver Chemistries Can Be Associated with Severe Ductal Disease and Dominant Strictures. Dig Dis Sci 2014:229–30.

[29] Olsson R, Hägerstrand I, Broomé U, Danielsson a, Järnerot G, Lööf L, et al. Sampling variability of percutaneous liver biopsy in primary sclerosing cholangitis. J Clin Pathol 1995;48:933–5.

[30] Angulo P, Larson DR, Therneau TM, LaRusso NF, Batts KP, Lindor KD. Time course of histological progression in primary sclerosing cholangitis. Am J Gastroenterol 1999;94:3310–3.

[31] Kawada T. The appropriate number of endpoints to keep validity for Cox proportional hazard analysis. Int J Cardiol 2011;153:110–1.

[32] Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. J Clin Epidemiol 1995;48:1495–501.

[33] Peduzzi P, Concato J, Feinstein, ARHolford T. Importance of eventers per independant variable in proportional hazard regression analysis II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;48:1503–10.

[34] Guido M, Rugge M. Liver biopsy sampling in chronic viral hepatitis. Semin Liver Dis 2004;24:89–97.

Figure legends

Fig.1. Distribution of grading and staging systems.

(A) Ishak grading system. (B) Nakanuma grading system.

(C) Ishak staging system. (D) Nakanuma staging system. (E) Ludwig staging system.

Fig. 2. Kaplan Meier survival curves. Endpoints transplant-free survival and time to liver transplantation, shown for: (A,B) Ishak staging system. (C,D) Nakanuma staging system. (E,F) Ludwig staging system.

Table 1. Patient characteristics.

N	64	
Male [n (%)]	40	(63)
Age follow-up (years) [mean (SD)]	49	(±15)
Age at diagnosis PSC (years) [mean (SD)]	38	(±14)
Large duct PSC [n (%)]	54	(84)
Inflammatory bowel disease [n (%)]	43	(67)
Ulcerative colitis [n (%)]	32	(50)
Crohn's disease [n (%)]	8	(12)
Unspecified [n (%)]	3	(5)
Portal tracts [median (IQR)]	13	(9-19)
Biopsy length (mm) [median (IQR)]	14	(11-19)
Disease duration at time of biopsy (months) [median (range)]	0	(0-20)
Follow up time (months) [median (IQR)]	112	(70-172)
AST xULN [median (IQR)]*	1.40	(1.04-2.64)
ALT xULN [median (IQR)]*	2.04	(1.40-4.43)
ALP xULN [median (IQR)]*	1.65	(1.24-3.39)
γGT xULN [median (IQR)]*	5.80	(3.02-11.10)
Total bilirubin xULN [median (IQR)]*	0.82	(0.52-1.21)
MRS*	-0.28	(-0.77-0.78)

SD = standard deviation; PSC = primary sclerosing cholangitis; IQR = inter quartile range; mm = millimeter; AST = aspartate aminotransferase; ALT = alanine aminotransferase; ALP = alkaline phosphatase; γGT = gamma-glutamyl transferase; xULN = times upper limit of normal; MRS = Mayo risk score
* Data available for: AST n=56, ALT n= 51, ALP n=52, γGT n=46 total bilirubin n= 49, MRS n=33.

Table 2. Prognostic significances of clinical, biochemical and histopathological parameters in predicting transplant-free survival, time to LTx and cirrhosis related symptoms, calculated by univariable Cox proportional hazard analyses.

	Transplant-free survival		**Time to LTx**		**Cirrhosis related symptoms***	
	HR (95% CI)	***p*-value**	**HR (95% CI)**	***p*-value**	**HR (95% CI)**	***p*-value**
Sex	0.97 (0.28 – 3.32)	0.96	0.57 (0.12-2.85)	0.50	0.37 (0.11-1.30)	0.12
Age at PSC diagnosis	1.01 (0.96 – 1.05)	0.63	1.00 (0.95-1.05)	0.97	1.01 (0.97-1.04)	0.70
PSC type	0.67 (0.09-5.28)	0.71	0.04 (0.00-301.42)	0.48	0.37 (0.05-2.78)	0.33
Coexisting IBD	0.43 (0.13-1.43)	0.17	0.66 (0.16-2.77)	0.57	1.95 (0.56-6.84)	0.30
Cirrhosis related symptoms*	1.19 (0.31-4.62)	0.80	2.07 (0.46–9.25)	0.34	NA	NA
AST*	1.18 (1.00-1.40)	0.05	1.11 (0.87-1.43)	0.41	1.07 (0.91-1.26)	0.42
ALT*	1.05 (0.91-1.21)	0.53	1.02 (0.84-1.24)	0.86	1.07 (0.97-1.19	0.18
ALP*	1.27 (0.96-1.68)	0.10	1.41 (1.03-1.93)	0.03	1.02 (0.77-1.34)	0.90
γGT*	0.95 (0.82-1.10)	0.51	0.98 (0.84-1.14)	0.78	1.05 (0.97-1.14)	0.20
Total bilirubin*	1.53 (0.99-2.35)	0.05	1.73 (1.07 – 2.79)	0.03	1.17 (0.80-1.70)	0.42
MRS*	3.41 (1.15-10.09	0.03	3.19 (0.97-10.47)	0.06	1.31 (0.72-2.38)	0.38
Ishak grade (1-4)	1.50 (0.86-2.61)	0.15	1.56 (0.82-2.99)	0.18	1.11 (0.69-1.79)	0.68
Nakanuma grade (0, 1, ≥2)	1.35 (0.65-2.82)	0.43	1.42 (0.60-3.31)	0.43	1.26 (0.69-2.30)	0.45
Ishak stage 1-3)	2.56 (1.11-5.89)	0.03	4.18 (1.51-11.56)	0.006	1.09 (0.95-3.81)	0.07
Nakanuma stage (1-4)*	6.53 (2.01-21.22)	0.002	7.05 (1.77-28.11)	0.006	1.57 (0.70-3.51)	0.28
Ludwig stage (1-4)	1.94 (1.00-3.79)	0.05	3.13 (1.42-6.87)	0.005	1.16 (0.66-2.00)	0.62

HR = Hazard Ratio; CI = confidence interval; LTx = liver transplantation; PSC = primary sclerosing cholangitis; IBD = inflammatory bowel disease; AST = aspartate aminotransferase; ALT = alanine aminotransferase; ALP = alkaline phosphatase; γGT = gamma-glutamyl transferase; MRS = Mayo risk score; NA = not assessed.
* Data available for: AST n=56, ALT n= 51, ALP n=52, γGT n=46, total bilirubin n= 49, MRS n=33, Nakanuma stage n=58; cirrhosis related symptoms n=63.

Table 3a. Correlation between grading and liver biochemistry values.

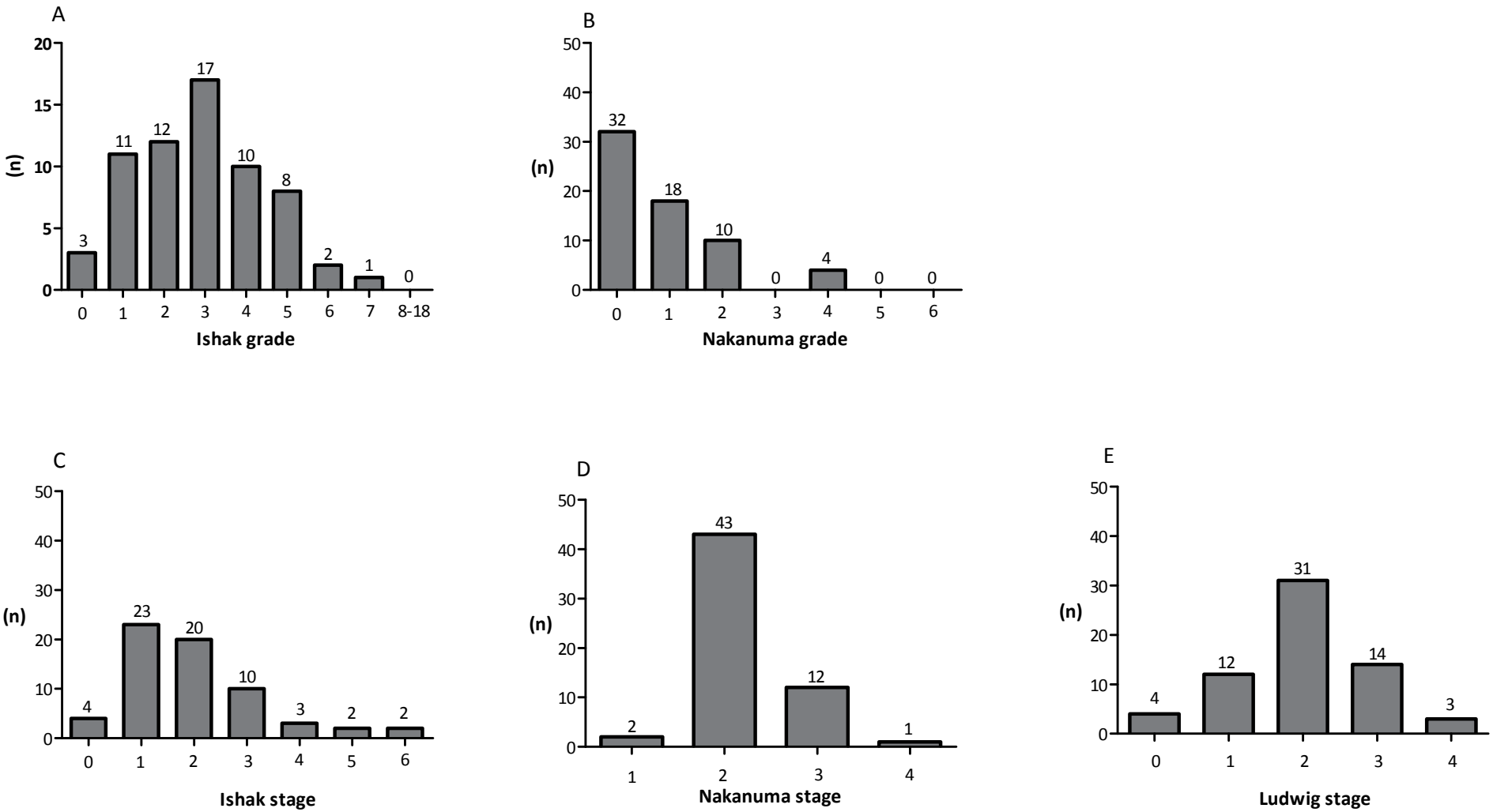
	Interface activity (0, 1, ≥2)	Confluent necrosis (0, ≥1)	Lobular inflammation (0, 1, ≥2)	Portal inflammation (0, 1, ≥2)	Ishak total (1-4)	CA (0, 1, ≥2)	HA (0, ≥1)	Nakanuma total (0, 1, ≥2)
AST*	0.062 p=0.650	-0.018 p=0.896	0.094 p=0.490	0.066 p=0.627	0.171 p=0.207	-0.038 p=0.783	0.236 p=0.080	0.100 p=0.462
ALT*	-0.122 p=0.393	-0.041 p=0.774	0.055 p=0.704	-0.025 p=0.863	-0.009 p=0.951	-0.201 p=0.158	0.052 p=0.716	-0.146 p=0.307
ALP*	0.256 p=0.067	-0.247 p=0.078	0.163 p=0.248	0.306 p=0.027	0.314 p=0.023	0.143 p=0.311	0.277 p=0.047	-0.132 p=0.360
γGT*	0.146 p=0.333	0.062 p= 0.683	-0.093 p=0.540	0.001 p=0.996	0.115 p=0.446	-0.018 p=0.905	0.239 p=0.110	0.126 p=0.402
Total* bilirubin	0.141 p=0.334	-0.011 p=0.940	-0.072 p=0.621	0.268 p=0.063	0.254 p=0.090	-0.241 p=0.095	0.093 p=0.526	-0.050 p=0.731

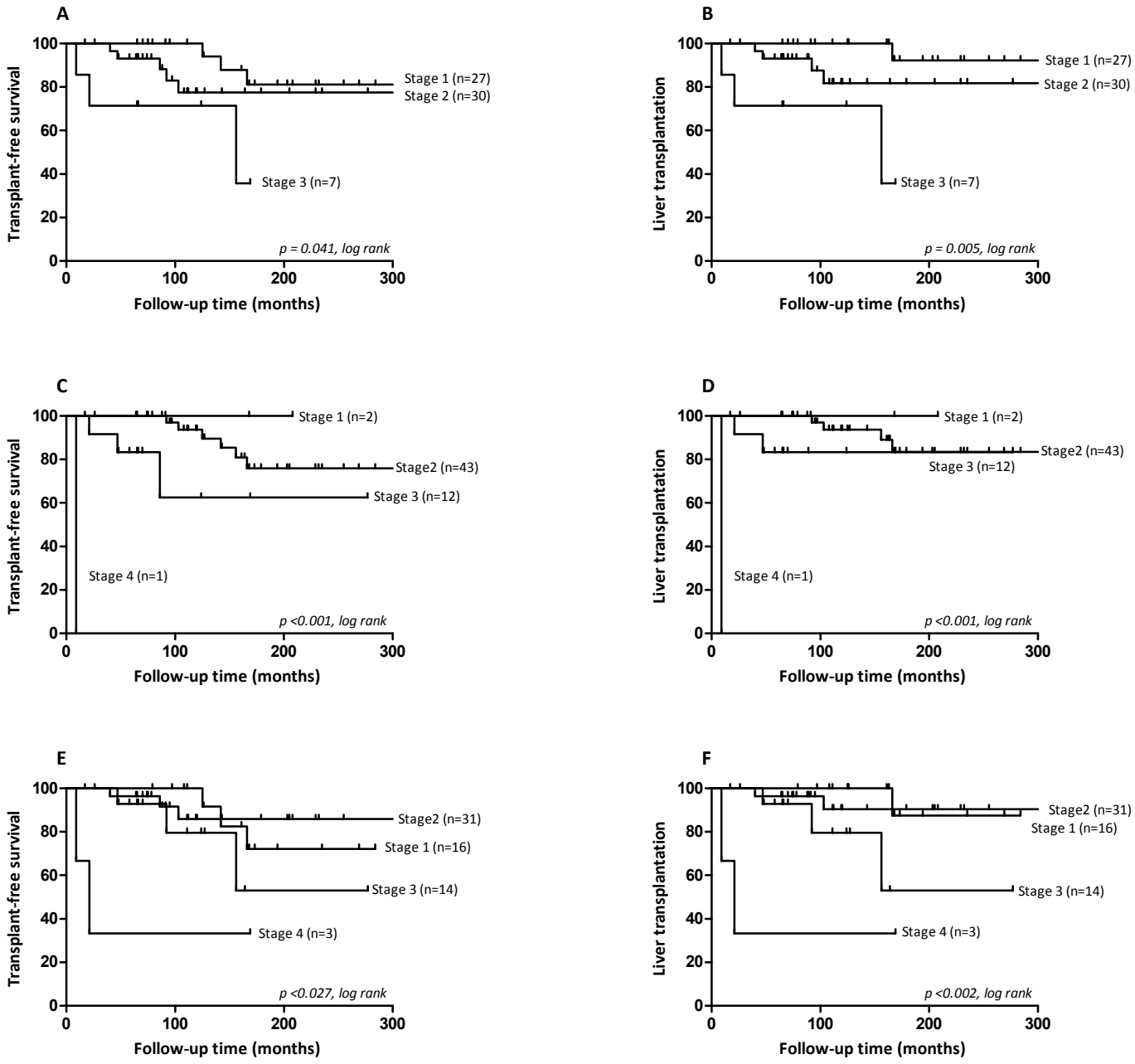
CA = cholangitis activity; HA = hepatitis activity; AST = aspartate aminotransferase; ALT = alanine aminotransferase; ALP = alkaline phosphatase; γGT = gamma-glutamyl transferase.
* Data available for: AST n=56, ALT n= 51, ALP n=52, γGT n=46, total bilirubin n= 49.

Table 3b. Correlation between staging and liver biochemistry value, calculated by Spearman's rank correlation test.

	Ishak (1-3)	Fibrosis (0-3)	Bile duct loss (0-3)	Orcein positive granules** (0-3)	Nakanuma total** (1-4)	Ludwig (1-4)
AST*	0.050 p=0.717	-0.028 p=0.840	0.039 p=0.778	-0.028 p=0.843	0.074 p=0.606	-0.011 p=0.938
ALT*	-0.167 p=0.241	-0.191 p=0.179	-0.040 p=0.780	-0.109 p=0.465	-0.085 p=0.569	-0.162 p=0.256
ALP*	0.180 p=0.201	0.088 p=0.535	0.001 p=0.992	0.024 p=0.870	0.146 p=0.322	0.131 p=0.356
γGT*	0.096 p=0.528	0.049 p=0.748	0.110 p=0.468	0.188 p=0.233	0.114 p=0.473	0.045 p=0.769
Total* bilirubin	0.317 p=0.027	0.292 p=0.042	0.245 p=0.090	0.121 p=0.429	0.255 p=0.091	0.090 p=0.537

AST = aspartate aminotransferase; ALT = alanine aminotransferase; ALP = alkaline phosphatase; γGT = gamma glutamyl transferase.
* Data available for: AST n=56, ALT n= 51, ALP n=52, γGT n=46, total bilirubin n= 49.
** Data available for: n=58.


A
(n)
20
15
10
5
0
3
11
12
17
10
8
2
1
0
0 1 2 3 4 5 6 7 8-18
Ishak grade
B
(n)
50
40
30
20
10
0
32
18
10
0
4
0
0
0 1 2 3 4 5 6
Nakanuma grade
C
(n)
50
40
30
20
10
0
4
23
20
10
3
2
2
0 1 2 3 4 5 6
Ishak stage
D
(n)
50
40
30
20
10
0
2
43
12
1
1 2 3 4
Nakanuma stage
E
(n)
50
40
30
20
10
0
4
12
31
14
3
0 1 2 3 4
Ludwig stage


A
Transplant-free survival
Follow-up time (months)
Stage 1 (n=27)
Stage 2 (n=30)
Stage 3 (n=7)
p = 0.041, log rank
B
Liver transplantation
Follow-up time (months)
Stage 1 (n=27)
Stage 2 (n=30)
Stage 3 (n=7)
p = 0.005, log rank
C
Transplant-free survival
Follow-up time (months)
Stage 1 (n=2)
Stage2 (n=43)
Stage 3 (n=12)
Stage 4 (n=1)
p <0.001, log rank
D
Liver transplantation
Follow-up time (months)
Stage 1 (n=2)
Stage2 (n=43)
Stage 3 (n=12)
Stage 4 (n=1)
p <0.001, log rank
E
Transplant-free survival
Follow-up time (months)
Stage2 (n=31)
Stage 1 (n=16)
Stage 3 (n=14)
Stage 4 (n=3)
p <0.027, log rank
F
Liver transplantation
Follow-up time (months)
Stage2 (n=31)
Stage 1 (n=16)
Stage 3 (n=14)
Stage 4 (n=3)
p <0.002, log rank